Environmental Health Perspectives Supplements
Vol. 102 Suppl 1:121-125 1994

Biostatistical Analysis of the Micronucleus Mutagenicity Assay Based on the Assumption of a Mixing Distribution

by Ludwig Hothorn

The *in vivo* micronucleus assay can be analyzed by comparing the number of micronuclei (MN) of several dose groups with those of a control group. In several publications, difficulties arose in estimating a suitable distribution for MN, even in the untreated historical control groups. Mitchell et al. described the presence of a subpopulation of more susceptible responders. Based on this assumption of such a subpopulation, score tests were used for the mixing distribution of responders and nonresponders (behavior same as in untreated control animals) within the dose groups. The power behavior of these tests was characterized with a simulation study. The advantage of score tests can be shown, even in the practical and important guideline case of only five animals per group.

Introduction

The statistical analysis of the *in vivo* micronucleus assay is based on significance tests for the differences between the numbers of micronucleated polychromatic erythrocytes (MN) in the control group and several dose groups. In several publications, difficulties arose in estimating a suitable distribution of the MN, even for the untreated case of historical control groups: *a*) Amphelett and Delow (*1*) described the validity of the Poisson distribution, *b*) Hart and Engberg-Petersen (*2*) found a good approximation to the binomial distribution, *c*) Mitchell et al. (*3*) reported a negative binomial distribution, *d*) Mackey and MacGregor (*4*) established an extra-binomial variation under treatment with clastogenic agents, *e*) Salsburg and Holden (*5*) detailed problems in choosing a suitable distribution for historical control data.

Mitchell et al. (*3*) discussed the presence of outliers in MN data in terms of a possible existence of a subpopulation of more susceptible responders. With this model, Ashby and Mirkova (*6*) explained the variation in the MN data. A theoretical background can be derived from the genetically based polymorphism in mammalian P-450 xenobiotic metabolizing enzymes (*7*). Another explanation is based on heritable strain differences in MN induced by polycyclic aromatic hydrocarbons (*8*). In addition, nonresponders may arise due to improper administration of the test substance in a single animal. This case will, however, not be considered here. Mitchell et al. (*3*) focused on an outlier analysis of historical control data in relation to concurrent control data and elimination of outliers with traditional statistical methods.

Due to the unclear distribution behavior of the outcome variable, rank tests were commonly used in several papers. For example, Leimer et al. (*9*) described the application of the Fisher-Pitman permutation test on micronucleus assay data. On the other hand, MacGregor et al. (*10*) recommended the use of Armitage's (*11*) trend test assuming a binomial distribution for MN in relation to the global number of polychromatic erythrocytes. In this respect, rank or permutation tests avoid the pooling of MN within the groups under the binomial sampling assumption and consider the importance of animal-to-animal variation. For this reason, special types of rank tests (so-called score tests), assuming a mixing distribution for the number of responders and nonresponders in the dose groups, will be considered here.

Analysis Based on the Mixing Distribution Assumption

Several methods assuming a mixing distribution can be used to solve the test problem. Here, only nonparametric score tests for the Lehmann (*12*) alternative hypothesis will be used (formulated as a one-sided, two-sample problem without limiting generalization).

Let $X_1, \ldots, X_m$ be the MN responses of the control group with the distribution function $H(x)$, and let $Y_1, \ldots, Y_n$, be the MN responses of a dose group with the distribution function $G(x)$. The hypotheses can be formulated under the mixing distribution assumption of responders and nonresponders in the dose group as:

Department of Biostatistics, German Cancer Research Center, Im Neuenheimer Feld 280, D-6900 Heidelberg, Germany.

This paper was presented at the International Biostatistics Conference on the Study of Toxicology that was held May 13-25, 1991, in Tokyo, Japan.

$$H_0 : H(x) = G(x)$$

$$H_A : H(x) < G(x) \text{ with } G(x) = pH(x) + (1-p)F_{patho}(x)$$

where p is the proportion of nonresponders, $(1\text{-}p)$ is the proportion of responders and p is assumed unknown.

Two types of Lehmann alternative will be considered here: shift

- $F_{patho}(x) = G(x - \delta)$

according to Good (*13*) and power

- $F_{patho}(x) = G^a(x)$

according to Lehmann (*12*). Johnson et al. (*14*) suggested approximate score statistics for the shift alternative based on the following mixed normal score function:

$$sm(i) = exp(-d^2/2)exp(d\Phi^{-1}(i/(m+n+1))^{-1}$$

where d is a constant (in the simulation study reported below, d=0.5,1,1.5,2 were used; here, only the case d=1 will be reported) and Φ is the distribution function of the standard normal distribution.

Conover and Salsburg (*15*) proposed the following approximate score function for the power alternative, as a generalizaton of Wilcoxon-Mann-Whitney (WMW) scores:

$$sc(i) = (i/(m+n+1))^{a-1}$$

where i is the rank in the combined $(x+y)$ sample, a is an integer constant (in the simulation study a=3,4,5,6 were used; here, only the case a=4 will be reported. In toxicology, tests based on this mixing distribution assumption have ben used for behavioral studies (*16*), teratological studies (*17*), sister chromatid exchange (*1*), and chronic studies (*18*).

Simulation Study

In a simulation study, two questions will be addressed: *a*) Is the assumption of such a mixing distribution a suitable approach for analyzing data from the micronucleus assay? *b*) Can we observe an increase in power (e.g., in relation to the commonly used WMW *U*-test), even in the guideline case considered here where n_j only equals 5?

The empirical distribution shown in Table 1 [(*3*) which approximates negative binomial distribution] was generated for the control groups using a PC program. Power estimations (based on 1000 replications) of the asymptotic two-sample test based on mixed normal scores [$sm(i)(d{=}1)$], asymptotic two-sample test based on generalized MWM-score [$sc(i)\ (a{=}4)$], and the WMW *U*-test were compared for a shift alternative (mean shift between the control and dose group) with shift parameters of {1,2,3}; standard deviations: s_c=1, s_D={1,2,3};α={0.01, 0.05, 0.10}; number of animals, n_j={5,10} and proportion of nonresponders, p: {0,0.2, 0.4,0.6,0.8}. In Table 2, the power estimations of the three tests under investigations are given for n_j=5. Under the null hypothesis, the α-estimations are quite close to the nominal levels, e.g., $\alpha = 0.10$ WMW:$\hat{\alpha} = 0.091$; $sm_{(i):}\alpha = 0.092$; $sc(i)$: $\hat{\alpha} = 0.014$.

Table 1. Empirical distribution of MN.

MN	Probability of MN/100 PCE
0	0.462
1	0.325
2	0.145
3	0.049
4	0.018
$\geq$4	0.001

Abbreviations: MN, micronuclei; PCE, polychromatic erythrocytes.

Table 2 shows that the score tests give a higher power than the WMW test for a medium-size effect between the control group and the dose group (represented by a shift 2) and the typical α level of 0.05, even for only one nonresponder in five animals (p=0.2). These power differences are not relevant for smaller shift parameters (e.g., 1). The differences arise with a larger α level, so that $n_j = 5$ and $\alpha = 0.01$, should be avoided.

The question that arises is whether increasing the number of animals up to 10 will give clear advantages of the score tests. Table 3 presents the related power estimations. For small and medium shift parameters, the increase in power of the score tests is higher in relation to the small sample size situation.

The power behavior dependent on the proportion of nonresponders p is given in Table 4. Table 4 presents the differences between the score tests and the WMW test. These are seen to be negligible both in the direction of a small proportion of nonresponders (unimodal distribution of all animals exhibiting a large reaction) and in the direction of a high proportion of nonresponders (unimodal distribution of animals exhibiting a small reaction; the estimation of p=0 equal to $\hat{\alpha}$ is not given in this table). Advantages of score tests are seen for proportions of p=0.2–0.8, whereby the dependence [based on the efficiency

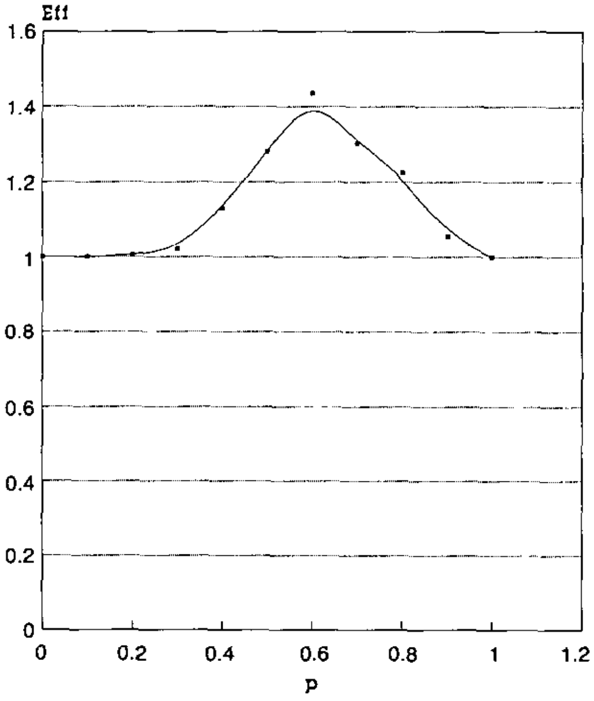

FIGURE 1. Efficiency criteria (*19*) on the proportion of nonresponders.

Table 2. Power of selected score tests for $n_j=5$ ($s_c=1$).

			Power estimations								
			$\alpha=0.10$			$\alpha=0.05$			$\alpha=0.01$		
s_D	Shift	p	WMW	Score $sc(i)$	Score $sm(i)$	WMW	Score $sc(i)$	Score $sm(i)$	WMW	Score $sc(i)$	Score $sm(i)$
1	1	0	0.67	0.59	0.61	0.51	0.45	0.47	0.28	0.22	0.23
1	1	0.2	0.51	0.48	0.49	0.34	0.33	0.33	0.15	0.12	0.12
1	1	0.4	0.34	0.37	0.37	0.22	0.22	0.21	0.07	0.06	0.06
1	1	0.6	0.25	0.26	0.26	0.14	0.14	0.14	0.05	0.04	0.04
1	1	0.8	0.16	0.18	0.18	0.09	0.09	0.09	0.03	0.02	0.02
2	1	0	0.75	0.77	0.75	0.63	0.57	0.63	0.35	0.30	0.30
2	1	0.2	0.59	0.59	0.69	0.42	0.41	0.40	0.17	0.15	0.13
2	1	0.4	0.40	0.45	0.44	0.26	0.25	0.26	0.08	0.07	0.07
2	1	0.6	0.27	0.30	0.30	0.15	0.16	0.15	0.05	0.04	0.04
2	1	0.8	0.17	0.19	0.19	0.10	0.09	0.09	0.03	0.02	0.02
3	1	0	0.81	0.78	0.82	0.67	0.63	0.68	0.37	0.33	0.31
3	1	0.2	0.61	0.62	0.65	0.44	0.43	0.46	0.19	0.16	0.15
3	1	0.4	0.41	0.47	0.49	0.26	0.27	0.27	0.08	0.07	0.07
3	1	0.6	0.27	0.31	0.31	0.16	0.16	0.16	0.05	0.04	0.02
3	1	0.8	0.17	0.19	0.19	0.10	0.09	0.09	0.03	0.02	0.02
1	2	0	0.96	0.91	0.92	0.89	0.83	0.84	0.69	0.68	0.68
1	2	0.2	0.81	0.82	0.81	0.65	0.70	0.68	0.30	0.30	0.28
1	2	0.4	0.56	0.65	0.63	0.39	0.42	0.40	0.13	0.12	0.11
1	2	0.6	0.33	0.42	0.39	0.20	0.21	0.21	0.06	0.05	0.05
1	2	0.8	0.19	0.23	0.22	0.11	0.11	0.11	0.03	0.03	0.03
2	2	0	0.98	0.98	0.97	0.95	0.89	0.92	0.75	0.72	0.72
2	2	0.2	0.86	0.87	0.87	0.69	0.75	0.72	0.32	0.31	0.28
2	2	0.4	0.58	0.68	0.67	0.39	0.44	0.43	0.13	0.12	0.11
2	2	0.6	0.34	0.44	0.40	0.20	0.22	0.21	0.06	0.05	0.05
2	2	0.8	0.19	0.24	0.22	0.11	0.11	0.11	0.03	0.03	0.03
3	2	0	0.98	0.96	0.97	0.96	0.91	0.93	0.77	0.73	0.74
3	2	0.2	0.87	0.89	0.89	0.71	0.75	0.74	0.33	0.32	0.28
3	2	0.4	0.58	0.69	0.63	0.40	0.44	0.43	0.13	0.12	0.11
3	2	0.6	0.34	0.41	0.40	0.20	0.22	0.21	0.06	0.05	0.05
3	2	0.8	0.19	0.23	0.22	0.11	0.11	0.11	0.03	0.03	0.03

WMW, Wilcoxon-Mann-Whitney score.

Table 3. Power selected score tests for $n_j=10$ ($s_c=1$).

			Power estimations								
			$\alpha=0.10$			$\alpha=0.05$			$\alpha=0.01$		
S_D	Shift	p	WMW	Score $sc(i)$	Score $sm(i)$	WMW	Score $sc(i)$	Score $sm(i)$	WMW	Score $sc(i)$	Score $sm(i)$
1	1	0	0.87	0.75	0.77	0.75	0.64	0.66	0.51	0.39	0.41
1	1	0.2	0.71	0.64	0.65	0.56	0.49	0.49	0.29	0.24	0.23
1	1	0.4	0.50	0.49	0.48	0.35	0.34	0.34	0.15	0.12	0.12
1	1	0.6	0.32	0.32	0.33	0.21	0.20	0.20	0.07	0.06	0.06
1	1	0.8	0.19	0.19	0.19	0.11	0.11	0.11	0.03	0.03	0.03
2	1	0	0.95	0.91	0.94	0.88	0.82	0.87	0.66	0.58	0.63
2	1	0.2	0.82	0.80	0.85	0.69	0.66	0.70	0.37	0.37	0.37
2	1	0.4	0.59	0.62	0.66	0.42	0.45	0.48	0.19	0.18	0.18
2	1	0.6	0.37	0.43	0.45	0.24	0.26	0.25	0.07	0.07	0.06
2	1	0.8	0.21	0.23	0.23	0.12	0.12	0.12	0.03	0.03	0.03
3	1	0	0.96	0.94	0.97	0.91	0.87	0.92	0.71	0.65	0.71
3	1	0.2	0.85	0.85	0.89	0.74	0.74	0.77	0.41	0.41	0.42
3	1	0.4	0.62	0.71	0.76	0.44	0.51	0.54	0.20	0.20	0.20
3	1	0.6	0.37	0.46	0.52	0.25	0.28	0.28	0.08	0.07	0.07
3	1	0.8	0.21	0.23	0.24	0.12	0.13	0.13	0.03	0.03	0.03
1	2	0	0.99	0.99	0.99	0.99	0.98	0.98	0.97	0.91	0.91
1	2	0.2	0.98	0.97	0.97	0.95	0.93	0.91	0.76	0.73	0.69
1	2	0.4	0.84	0.87	0.84	0.68	0.72	0.70	0.34	0.39	0.34
1	2	0.6	0.50	0.61	0.60	0.34	0.44	0.41	0.11	0.12	0.11
1	2	0.8	0.24	0.30	0.30	0.13	0.16	0.15	0.04	0.04	0.03
2	2	0	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.96	0.98
2	2	0.2	0.99	0.99	0.99	0.97	0.97	0.98	0.81	0.83	0.80
2	2	4	0.87	0.92	0.91	0.71	0.81	0.79	0.36	0.44	0.38
2	2	0.6	0.51	0.68	0.70	0.35	0.48	0.45	0.11	0.13	0.11
2	2	0.8	0.25	0.31	0.32	0.13	0.16	0.16	0.04	0.04	0.03
3	2	0	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.98
3	2	0.2	0.99	0.99	0.99	0.98	0.97	0.98	0.82	0.85	0.82
3	2	0.4	0.88	0.93	0.93	0.72	0.83	0.82	0.37	0.45	0.40
3	2	0.6	0.51	0.71	0.73	0.36	0.49	0.47	0.11	0.13	0.11
3	2	0.8	0.25	0.32	0.33	0.14	0.17	0.16	0.04	0.04	0.03

WMW, Wilcoxon-Mann-Whitney score.

Table 4. Power $= f$(proportion of nonresponders) ($n_j=10$, $s_D=2$, shift$=2$).

	$\alpha=0.10$			$\alpha=0.05$			$\alpha=0.01$		
p	WMW	Score $sc(i)$	Score $sm(i)$	WMW	Score $sc(i)$	Score $sm(i)$	WMW	Score $sc(i)$	Score $sm(i)$
0	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.96	0.98
0.1	0.99	0.99	0.99	0.99	0.99	0.99	0.95	0.93	0.93
0.2	0.99	0.99	0.99	0.97	0.97	0.98	0.81	0.83	0.80
0.3	0.96	0.97	0.97	0.89	0.91	0.91	0.57	0.64	0.59
0.4	0.87	0.91	0.91	0.71	0.81	0.79	0.36	0.44	0.38
0.5	0.71	0.83	0.84	0.52	0.66	0.63	0.23	0.26	0.24
0.6	0.51	0.68	0.70	0.35	0.48	0.45	0.11	0.13	0.11
0.7	0.36	0.50	0.50	0.22	0.29	0.28	0.07	0.06	0.06
0.8	0.25	0.31	0.32	0.13	0.16	0.16	0.04	0.04	0.03
0.9	0.17	0.18	0.18	0.09	0.10	0.10	0.02	0.02	0.02

WMW, Wilcoxon-Mann-Whitney score.

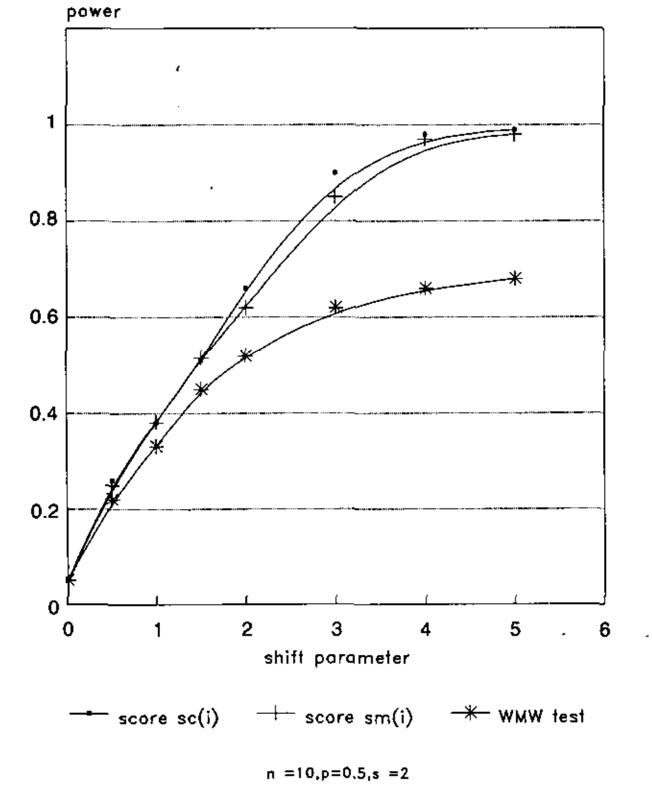

FIGURE 2. Power function of the WMW-test and score tests.

measure according to Lee and Wolfe (*19*)] is not symmetric at about $p=0.5$ (Fig. 1).

Only one point of the power function is given in Tables 2–4. Therefore, the power functions for selected values of p,n_j,s_D and α are shown in Figure 2. For medium-size shifts, the differences among the power functions are important up to a maximum shift value (decreasing with smaller α levels), after which parallelism of the power functions holds true.

These simulation results for the biostatistical analysis of the micronucleus assay suggest that the score tests have an advantage in power in relation to the commonly used WMW test. These advantages are particularly relevant for *a*) medium to large effect differences between the control and dose group, *b*) ranges of $p \approx 0.2,\ldots,0.8$, *c*) values of $n_j = 5$ and $\alpha = 0.05$. This advantage increases as the sample size, n_j, and α level become larger.

Table 5. Experimental MN data.

Dose	MN	Pooled data
Control	3 2 1 1 3 2 0 3 2 1 3 1 2 1 3 3	31/16000
5	2 2 1 0 1 2 0 1	9/8000
10	3 3 6 1 4 3 3 1	24/8000
20	3 7 3 8 4 6 7 4	42/8000
40	26 25 23 34 25 23 19 29	204/8000

MN, miconuclei

Table 6. Statistical test results.

	Dose groups			
	5	10	20	40
Asymptotic WMW test				
p-Value	0.97	0.08	0.002	0.001
Significance $\alpha=0.05$	—	—	<	<
Exact Fisher permutation test				
p-Value	0.98	0.047	0.001	0.001
Significance $\alpha=0.05$	—	<	<	<
Asymptotic $sc(i)$ score test				
p-Value	0.98	0.025	0.001	0.001
Significance $\alpha=0.05$	—	<	<	<

WMW, Wilcoxon-Mann-Whitney score; <, significant increase.

The micronucleus assay sometimes represents a "control versus k dose groups design" for a one-sided, ordered alternative hypothesis (because only increasing MNs with increasing doses are considered biologically significant). Based on the two-sample tests described above, a simple *a priori* ordering procedure (*20*) can be used in this case.

An Example

Experimental data from Kliesch et al. (*21*) were used for a micronucleus assay on mice, 24 hr after single *per os* treatment of methyl methane sulfonate (MMS) (Table 5). Results of the biostatistical analysis are shown in Table 6. This example shows the greater sensitivity for the contrast between the control group and dose group 10 for both the Fisher permutation test and the score test.

Conclusions

From the results presented here, one can conclude that the choice of statistical method for the analysis of micronucleus assay data when MN is increasing relative to controls is not critical at the commonly used level of $\alpha=0.05$. However, a suitable choice of test is necessary for small or medium-sized increases in numbers of MN. This is applicable, for example, in the case of the no-observed-effect dose estimation. With a simulation study, based on an empirical negative binomial distribution of MN and a shift alternative, an advantage in the power behavior of selected score tests assuming a mixing distribution of responders and non-responders is evident, even for the guideline case $n_j=5$, $\alpha\geq0.05$, $p\geq0.1$, and a medium-sized shift between dose and control groups.

REFERENCES

1. Amphlett, G. E., and Delow, G. F. Statistical analysis of the micronucleus test. Mutat. Res. 128: 161–166 (1984).
2. Hart, J. W., and Engberg-Petersen, H. Statistics of the mouse bone-marrow micronucleus test: counting, distribution and evaluation of results. Mutat. Res. 111: 195–207 (1983).
3. Mitchell, I. G., Carlton, J. B., and Gilbert, P. J. The detection and importance of 'outliers' in the in vivo micronucleus assay. Mutagenesis 3: 491–495 (1988).
4. Mackey, B. E., and MacGregor, J. T. The micronucleus test: statistical design and analysis. Mutat. Res. 64: 195–204 (1979).
5. Salsburg, D., and Holden, E. A statistical examination of historical controls for mouse bone marrow cytogenetic assays. Environ. Mutagen. 4: 55–62 (1985).
6. Ashby, J., and Mirkova, E. The activity of MNNG in the mouse bone marrow micronucleus assay. Mutagenesis 2: 199–204 (1987).
7. Rampersaud, A., and Walz, F. G. Polymorphism of four hepatic cytochromes P-450 in twenty-eight inbred strains of rat. Biochem. Genet. 25: 527–534 (1987).
8. Sato, S., Kitajima, H., Takizawa, H., and Inui, N. Mouse strain differences in the induction of micronuclei by polycyclic aromatic hydrocarbons. Mutat. Res. 192: 185–189 (1987).
9. Leimer, I., Peil, H., and Ellenberger, J. Statistical analysis of the micronucleus test with the Fisher-Pitman permutation test. In: Statistical Methods in Toxicology. Lecture Notes in Medical Informatics, Vol. 43 (L. Hothorn, Ed.), Springer-Verlag, Heidelberg, 1991, pp. 20–24.
10. MacGregor, J. T., Heddle, J. A., Hite, M., Margolin, B. H., Ramel, C., Salamone, M. F., Tice, R. R., and Wild, D. Guidelines for the conduct of micronucleus assays in mammalian bone marrow erythrocytes. Mutat. Res. 189: 103–112 (1987).
11. Armitage, P. Tests for linear trends in proportions and frequencies. Biometrics 11: 375–386 (1955).
12. Lehmann, E. L. The power of rank tests. Ann. Math. Stat. 24: 23–43 (1953).
13. Good, P. I. Detection of a treatment effect when not all experimental subjects will respond to treatment. Biometrics 35: 483–489 (1979).
14. Johnson, R. A., Verrill, S., and Moore, D. H. Two-sample rank tests for detecting changes that occur in a small proportion of the treated population. Biometrics 43: 641–655 (1987).
15. Conover, W. J., and Salsburg, D. S. Locally most powerful tests for detecting treatment effects when only a subset of patients can be expected to 'respond' to treatment. Biometrics 44: 189–196 (1988).
16. Nation, J. R. The effects of oral cadmium exposure on a passive avoidance performance. Toxicol. Lett. 20: 41–47 (1984).
17. Cory-Slechta, D. A. Chronic postweaning lead exposure and response duration performance. Toxicol. Appl. Pharmacol. 60: 78–84 (1981).
18. Hothorn, L. General principles in testing of toxicological studies. In: Statistical Methods in Toxicology. Lecture Notes in Medical Informatics, Vol. 43 (L. Hothorn, Ed.), Springer-Verlag, Heidelberg, 1991, pp. 111–131.
19. Lee, Y. J., and Wolfe D. A. A distribution-free test for stochastic ordering. J. Am. Stat. Assoc. 71: 722-727 (1976).
20. Hothorn, L., and Lehmacher, W. A simple testing procedure 'control versus k treatments' for one-sided ordered alternatives, with application in toxicology. Biometr. J. 33: 179–189 (1991).
21. Kliesch, U. I., Danford, N., and Adler, I.-D. Micronucleus test and bone-marrow chromosome analysis. A comparison of 2 methods in vivo for evaluating chemically induced chromosome alterations. Mutat. Res. 80: 321–332 (1981).